Original Article

Predicting the performance of batsmen in test cricket

INDIKA PRADEEP WICKRAMASINGHE

Eastern New Mexico University, United States

ABSTRACT

Wickramasinghe, I.P. (2014). Predicting the performance of batsmen in test cricket. *J. Hum. Sport Exerc., 9*(4), pp.744-751. Cricket is one of the team games played over 50 countries in different levels. Though the performance of each batsman in the team can be easily quantified, the prediction of player performance is arduous. This paper demonstrates a methodology to predict the performance of cricket batsman in test-match series. In this study, longitudinal test cricket data have been collected over five years of period. A model is developed to predict the player performance as a function of certain characteristics related to the player, the team and the match series. Due to the hierarchical nature of the collected cricket data, a three stage hierarchical linear model is proposed in this investigation. According to the outcome of the analysis, the handedness of the player (batsman) and the rank of the team significantly influence player performance. Finally, an accurate prediction of player performance is conducted using the proposed model.
Key words: CRICKET PLAYER, LONGITUDINAL, HIERARCHICAL LINEAR MODEL.

Corresponding author. Eastern New Mexico University ,1500 S Avenue K, Portales, NM 88130, United States.
E-mail: Indika.Wickramasinghe@enmu.edu
Submitted for publication September 2013
Accepted for publication May 2014
JOURNAL OF HUMAN SPORT & EXERCISE ISSN 1988-5202

doi:10.14198/jhse.2014.94.01

INTRODUCTION

Cricket is a bat and ball team game that comprises of three formats, namely, test cricket, one day cricket (ODI), and twenty20. The emphasis of this work is to test cricket, which is the longest form of cricket and is regarded by players and serious cricket fans as the ultimate test of playing ability. One of the main facets of test cricket is batting, the act or skill of hitting the cricket ball with a cricket bat to score runs. As with all other cricket statistics, batting statistics and records are a major part of the game and provide a measure of a player performance. Due to the popularity of the game and the rich statistics nature, predicting player performance has become a major importance.

A comprehensive review of the literature regarding the performance of both the player and the game reveals following findings. Stretch (2003) researched about cricketers' injuries that could affect their performances in the game. In this study, the author applied a hierarchical linear model (HLM) to model the nature of injuries to South African cricketers, including doctors and physiotherapists working with the South African team. Kimber and Hansford (1993) proposed a method, which was based on nonparametric approach to assess the batting performance of cricket batsmen. Reaction time is regarded as one of the incalculable talents of cricketers in all the departments of the game. Balasaheb, Maman, and Sandhu (2008) attempted to find the impact of visual skills training that could affect the performance of batsman. In their research, they showed how the visual skills improve the reaction time, depth perception, and eye of the cricketers, which eventually improves the batting performance of the player. While discussing about necessary adjusted measures to analyze the player performance in the game of cricket, Lemmer (2008) investigated the performances of players' when players participate in small number of cricket matches. Christie (2012) discussed the physical demands needed for the batting performance of cricketers'. In that work, the author pointed out that often cricketers do not pay adequate attention to the physical demand of the game, and suggested the necessity of good training sessions to improve the performance. In addition to these, Christie addressed the psychological and the musculoskeletal demands in cricket, which have an impact on the player performance. Saikia, Bhattacharjee, and Bhattacharjee (2012) compared the performance of both Indian and foreign cricketers who took part in Indian Premier League (IPL) twenty20 cricket tournament. They studied how the player performance changed when the player represented the IPL tournament and his national team. Ultimately, they proposed a model that considered characteristics such as the number of innings that a player played, the strike rate, and the batting average to measure the player performance. Nawaz, Mubeen, and Aleem (2012) used an individual and moving range control charts to evaluate the performance of cricketers. Lewis (2005) used the well established "Duckworth/Lewis" methodology to model an alternative measures to analyze player performance in the game of cricket. Brooks, Bussie`re, Jennions, and Hunt (2003) studied the importance of the handedness of the player on his performance. According to their findings, left-handed batsmen have performed better than the right-handed batsmen in the global event of cricket world cup in 2003. They extended their findings by demonstrating that most successful teams in the above world cup tournament had nearly 50% of left-handed players and it was a strategic advantage for their accomplishments. When the performance of cricketers' is concerned, the literature shows that the height of players is an important trait to consider. One of the variables of interest in the study of Houghton (2010) was the height of player, which could have a direct impact on scoring runs in a match. Koley et al. (2012) investigated the influence of variables related to anthropometric, physical strength, and body composition on the performance of cricketers and height was one of the variables under their consideration. Morley and Thomas (2005) investigated the influence of home advantage and other factors that influence the status of the match. In their investigation, Totterdell and Leach (2001) discussed how the experience of the player could increase the performances of professional sportsmen including cricketers. Singh et al. (2011) evaluated the batting performances of

cricketers and how it influences the annual rankings, which is published by the International Cricket Council (ICC).

HLM, also known as multi-level analysis is a longitudinal data analysis technique that can be applied to the data with a hierarchical structure. In this cricket data, runs scored by a player in a match series is nested within the player. Furthermore, the player is again nested within the team. In addition to these, the collected cricket data exhibits longitudinal nature due to the repeated observations (average runs scored in a match series) of the same player over a long periods of time. Therefore, the collected data exhibits both the longitudinal nature and the hierarchical structure with three levels. The development in the statistical theory of HLMs enables an integrated approach to study the structure of inter-individual and intra-individual growth and identify inter-individual predictors of developmental patterns (Raudenbush and Bryk, 2002). This idea of modeling inter-individual and intra-individual growth can be incorporated to model the player performance in test matches.

Therefore in this paper, a three-level HLM is used to model the performance of batsmen in the game of cricket. In this analysis, some anthropometric characteristics such as the height and the handedness of the player are considered. Team-level and cricket-match-level variables such as the rank of the team, the location of the match was played, and the match-series number that the player participated is included into this analysis. One of the main goals of this work is to find answers to the questions of what player-level and team-level characteristics influence the player performance. After identifying the variables that have an impact on the performance, a HLM is formed to predict the performance of batsmen.

MATERIAL AND METHODS

Sample

Test cricket batsmen, who played during the period of 2006-2010, were considered in this study. These players were selected from nine different countries, namely, Australia (Aus), Bangladesh (Ban), England (Eng), India (Ind), New Zealand (NZ), Pakistan (Pak), South Africa (SA), Sri Lanka (SL), and West Indies (WI). Above nine countries were the listed test cricket-playing nations, according to the ICC, during the data collection period. Cricket batsmen, who had played more than 10 test matches and averaged over 25 runs per game (or those who have scored a century or over 5 fifties) during their career were considered for the selection. In addition to the above criteria, batsmen who had played at least 3 test match series were considered. Average runs scored (RUNS), during each test match series were recorded along with the series number (S_NO) that the batsman played, and the place where each match series was played (HOME). If the series was played in batsman's country, 1 was assigned the variable HOME, and if it was away from his country, 0 was assigned. In each series, batsmen who played at least two matches were considered for this analysis. The rank of each team (RANK) was calculated by averaging their respective annual team ranks published by the ICC, the governing body of cricket. Whether the player is a left hander (HAND=0) or a right hander (HAND=1) was recorded using the variable HAND. Finally, the height (HEIGHT) of each batsman was recorded in meters.

Analysis

As explained before, this collected data exhibits both the longitudinal nature and the hierarchical structure of three levels. Therefore, the entire data can be represented according to three levels as follows. At level one, each individual's observed performance (average runs in a match series) is expressed as a function of some characteristics that are related to the match series and a random error component. At level two, it is assumed that the level one intercept is a function of certain measurable individual characteristics and a

random error term. Finally, at level three, level two intercept is expressed as a function of next level characteristics and random error terms.

- Level-1 Model
At level-1 the outcome variable RUNS, is expressed as follows.

$$RUNS_{ijk} = \pi_{0jk} + \pi_{1jk}*(S\_NO_{ijk}) + \pi_{2jk}*(HOME_{ijk}) + e_{ijk}$$

where, $RUNS_{ijk}$ is the average runs in the i^{th} series by the j^{th} batsman of k^{th} team, π_{0jk} is the intercept for the level 1, π_{1jk} is the regression coefficient associated with the *S_NO* of the j^{th} batsman of k^{th} team, π_{2jk} is the regression coefficient associated with the *HOME* of the j^{th} batsman of k^{th} team, S_NO_{ijk} is the i^{th} series of the j^{th} player of k^{th} team, $HOME_{ijk}$ is the place where the i^{th} series of the j^{th} batsman of k^{th} team, e_{ijk} is the level-1 random effect and σ^2 is the variance of random term e_{ijk}.

- Level-1 Model
In this level, intercept of the level one, π0jk is expressed as a function of HEIGHT, HAND and random component.

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}*(HEIGHT_{jk}) + \beta_{02k}*(HAND_{jk}) + r_{0jk}, \qquad \pi_{1jk} = \beta_{10k}, \qquad \pi_{2jk} = \beta_{20k}$$

Where, $HEIGHT_{jk}$ is the height of the j^{th} batsman of the k^{th} team, $HAND_{jk}$ is whether the j^{th} player of the k^{th} team a right hander or a left hander, β_{00k}, β_{01k}, β_{02k}, β_{10k} and β_{20k} are level-2 coefficients, r_{0jk} is the level-2 random term.

- Level-3 Model
In this level, intercept of level two, β_{00k} is considered as a function of the RANK and random component.

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(RANK_k) + u_{00k}, \quad \beta_{01k} = \gamma_{010}, \qquad \beta_{02k} = \gamma_{020}, \qquad \beta_{10k} = \gamma_{100}, \qquad \beta_{20k} = \gamma_{200}$$

where, $RANK_k$ is the rank of the k^{th} team, $\gamma_{000}, \gamma_{010}, \gamma_{020}, \gamma_{100}$ and γ_{200} are nonrandom effects u_{00k} is the level-3 random effect.

By combining each of the above three models the mixed-model can be expressed as follows.

- Mixed Model

$$RUNS_{ijk} = \gamma_{000} + \gamma_{001}*RANK_k + \gamma_{010}*HEIGHT_{jk} + \gamma_{020}*HAND_{jk} + \gamma_{100}*S\_NO_{ijk} + \gamma_{200}*HOME_{ijk} + r_{0jk} + u_{00k} + e_{ijk}$$

RESULTS

In this collected data set, the average number of test match series that batsmen have played during this time period was 6.30 (SD=4.62). The average number of match series played in batsman's home countries was nearly half (49%). The response variable of the study, the average runs scored by batsman in each match series was about 38.30 (SD=20.53). The average height of the batsman was 1.80 m (SD=0.09) and

65% of the batsmen were right handed batsmen. Australia, England, India, South Africa and Sri Lanka were the teams with a higher rank.

Tables 1, 2 and 3 show the estimated parameters for both random and non-random components. According to the table 1, RANK, γ_{001}= 4.223, t (7) = 2.736, p<0.05 and HAND, γ_{020} = -3.407, t (149) = -2.107, p<0.05 have significant impacts on the player performance. The significance of S_NO and HEIGHT are negligible. According to this results, player's handedness (HAND) negatively associates with RUNS. Though HOME positively influence the performance (RUNS), the impact is not statistically significant.

Table 1. *Estimation of Level-2 and Level-3 fixed effects*

Fixed Effect	Estimates	SE	t
Model for the RUNS means, π_0			
For Intercept 2, β_{00}			
Intercept 3, γ_{000}	51.523	14.574	3.535**
RANK, γ_{001}	4.223	1.544	2.736*
For HEIGHT, β_{01}			
Intercept 3, γ_{010}	-9.055	7.963	-1.137
For HAND, β_{02}			
Intercept 3, γ_{020}	-3.407	1.617	-2.107*
Model for S_NO slope, π_1			
Intercept 2, β_{10}			
Intercept 3, γ_{100}	0.119	0.091	1.317
Model for HOME/RUNS slope, π_2			
Intercept 2, β_{20}			
Intercept 3, γ_{200}	1.723	1.179	1.461

*p < .05; **p < .01.

Table 2. *Estimation of Level-2 and Level-3 random effects*

Random Effect	Variance component	df	χ^2
Level 1, e_{tij}	373.807		
Level 2, r_{0ij}	37.017	147	263.988*
Level 3, u_{00j}	0.006	7	7.722

*p < .05.

Table 3. *Variance Components for Random Effects at Levels 1, 2 and 3*

Level-1	$\hat{\sigma}^2$	= 373.807
Level-2	$\hat{\tau}_{\pi}$	= 37.017
Level-3	$\hat{\tau}_{\beta}$	= 0.006

In Table 4, prediction of player performance is presented using the fitted model. Sixteen batsmen (two players from each team) are displayed in the Table 4 to illustrate the prediction. Out of all the players, only two players, who had highest ICC ranks in each team and who played the immediate match series after the data collection period, were considered for table 4. A comparison is made between the average predicted runs and the actual average runs scored by above batsmen in the next immediate match series that each respective teams participated after the data collection period was over. This predicted runs were calculated using the proposed model. When the absolute error of the prediction is considered, it is apparent that this prediction works well for most of the players, but not for few players such as HM Amla (SA), Y Khan (Pak) and KC Sangakkara (SL).

Table 4. *Predicted runs vs. actual runs scored by players*

Player	Country	Predicted runs	Actual runs	Abs (error)
RT Ponting	Aus	47.33	40.09	7.24
MJ Clarke	Aus	42.61	42.80	0.19
Mushfiqur Rahim	Ban	37.14	36.50	0.79
Shakib Al Hasan	Ban	53.00	51.25	1.75
VVS Luxman	Ind	46.86	48.60	1.74
R Dravid	Ind	46.51	50.20	3.69
BB McCullum	NZ	38.43	39.25	0.82
LRPL Taylor	NZ	36.37	36.00	0.37
Taufeeq Umar	Pak	39.57	41.25	1.68
Younis Khan	Pak	43.36	59.00	15.64
AB de Villiers	SA	47.30	48.33	1.03
HM Amla	SA	48.27	59.75	11.48
DPMD Jayawardene	SL	50.53	54.00	3.47
KC Sangakkara	SL	45.51	30.66	14.85
S Chanderpaul	WI	44.55	48.20	3.65
MN Samuels	WI	35.95	36.00	0.05

DISCUSSION AND CONCLUSION

The relationship between the ranks of the team the player performance can be easily understood. The impact of negative association of the handedness confirms the finding of Brooks, Bussie`re, Jennions, and

Hunt (2003). This may be due to the fact that several left-handed batsmen in this collected data set have scored runs steadily throughout this time duration (2006-2010). MEK Hussey (Aus), AN Cook (Eng), G Gambhir (India), AC Gilchrist (Aus), GC Smith (SA), KC Sangakkara (SL), and S Chanderpaul (WI) were some of such players. The Height of the player also has become an insignificant variable as far as their performance is concerned. The main surprise of this analysis is the insignificant-positive- influence of the place of the match (HOME) on player performance. In the past, experts of the game of cricket, used to believe that batsmen are more familiar with their own conditions and expect to perform more in their home country than in series played outside of the batsman's country. There are several possible reasons for this unexpected result. Unlike in 1990s, present game of cricket has changed in various ways. With the emergence of twenty20, the shortest version of cricket, different countries started to organize several twenty20 tournaments around the world. Due to this, test cricket players found the opportunity to play in other conditions more often than early days. This not only helped batsmen to become more and more familiar with different conditions, but also to face world-famous bowlers regularly. When their national cricket teams tour on these countries, where the above mentioned tournaments are organized, these players get an advantage to score more runs than the batsmen of early days. Another possible reason for this could be the fact that during this period, there were only few matches played in Pakistan. Most of the matches were played in neutral counties due to security concerns in Pakistan. Though unexpected, the number of series that the batsman played does not have an impact on the performance. One reason behind this could be the emergence of twenty20 matches in various countries. According to the collected data, some players such as SE Marsh (Aus), EJG Morgan (Eng), RS Bopara (Eng), V Kohli (India), AD Mathews (SL), and KA Edwards (WI) have performed really well during their initial test match series. This could be another reason for the above unexpected findings. Due to the involvement of various variables, it is not an easy task to predict the performance of batsmen in a test match series with high accuracy. It is very interesting that this proposed model based on HLM, predicts reasonably well irrespective of some inconsistencies with a few batsmen.

Though it is interesting to consider additional attributes of both the batsman and the team for this analysis, it is not an easy task due to the increasing nature of the parameters. In addition, the proposed model takes care of both inter-individual and intra-individual characteristics of the player that could affect the performance. Approaches focusing on inter-individual variation emphasize the establishment of general developmental principles that apply to all individuals. In contrast, approaches focusing on intra-individual variation emphasize understanding the change within the individual (Collins, 2006). Cross-sectional data analysis can be regarded as an alternative approach for this model that uses a longitudinal approach. Longitudinal studies are in contrast to cross-sectional studies, in which a single outcome is measured for each individual where as in longitudinal study individuals are measured repeatedly through time. Though it is possible to address the same scientific questions with a longitudinal or a cross-sectional study, the major advantage of the former is its capacity to separate what in the context of population studies are called cohort and age effects (Diggle, Liang, and Zeger , 2002).

REFERENCES

1. Balasaheb, T., Maman, P., & Sandhu J. S. (2008). The impact of visual skills training on batting performance in cricketers. *Serbian Journal of Sports Sciences,* 2(1), pp.17-23.
2. Brooks, R., Bussie`re, L. F., Jennions, M. D., & Hunt, J. (2003). Sinister strategies succeed at the cricket World Cup. *Proceedings of the Royal Society*.
3. Morley, M., & Thomas, D. (2005). An investigation of home advantage and other factors affecting outcomes in English one-day cricket matches. *Journal of Sports Science,*23(3):261-268.

4. Christie, J. (2012). The Physical Demands of Batting and Fast Bowling in Cricket. *An International Perspective on Topics in Sports Medicine and Sports Injury*, Dr.Kenneth R. Zaslav (Ed.), ISBN: 978-953-51-0005-8.
5. Collins, L. M. (2006). Analysis of Longitudinal Data: The Integration of Theoretical Model, Temporal Design, and Statistical Model. *Annual Review of Psychology,* 57, pp.505-528.
6. Daniyal, M., Nawaz, T., Mubeen, I., & Aleem, M. (2012). Analysis of Batting Performance in Cricket using Individual and Moving Range (MR) Control Charts. *International Journal of Sports Science and Engineering*, *6*(4), pp.195-202.
7. Diggle, P. J., Liang, K., & Zege, S. L. (2002). Analysis of Longitudinal Data. *Oxford Science Publications.*
8. Kimber, A. C., & Hansford, A. (1993). A Statistical Analysis of Batting in Cricket. *Journal of the Royal Statistical* Society, *156*(3), pp.443-455.
9. Koley, S., Kumaar, B. S., & Shadagopan, S., P. (2012). Anthropometric, Physical Strength, Body Composition and Performance Test Profiles of Inter-District Level Male Cricketers of Punjab, India. *Anthropologist*, *14*(5), pp. 445-451.
10. Houghton, L. A. (2010). Running Between the Wickets in Cricket: What is the Fastest Technique? *International Journal of Sports Science & Coaching*, *5*(1), pp.101.
11. Lemmer, H. H. (2008). Analysis of players' performances in the first cricket twenty 20 world cup series. *South African Journal for Research in Sport*, *30*(2), pp.71-77.
12. Lewis, A. J. (2005). Towards fairer measures of player performance in one-day cricket. *Journal of the Operational Research Society, 56*, pp.804-815.
13. Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed.)*. Newbury Park, CA, Sage.
14. Raymond, M., Pontier, D., Dufour, A. B., & Møller, A. P. (1996). Frequency-dependent maintenance of left handedness in humans. *Proceedings of The Royal Society* B, *263*, pp.1627-1633.
15. Saikia, H., Bhattacharjee, D., & Bhattacharjee, A. (2012). Is IPL Responsible for Cricketers' Performance in Twenty20 World Cup? *International Journal of Sports Science and Engineering, 6*(2), pp.96-110.
16. Stretch, R. A. (2003). Cricket injuries: a longitudinal study of the nature of injuries to South African cricketers. British *Journal of* Sports Medicine, *37*, pp.250-253.
17. Swartza, T. B.*,* Gillb, P. S., Beaudoinc, D., & De Silv, B. M. (2006). Optimal batting orders in one-day cricket. *Computers & Operations Research, 33*, pp.939-1950.
18. Totterdell, P., & Leach, D. (2001). Negative mood regulation expectancies and sports performance: an investigation involving professional cricketers. *Psychology of Sport and Exercise, 4*(2), pp.249-265.

